知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum
E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于“拉开档次”法的网络信息污染状况模糊语言评价研究

◎ 万晓榆 王在宇 蒋婷

重庆邮电大学经济管理学院 重庆 400065

摘要：[目的 / 意义] 网络信息污染已经为广大人民群众所深恶痛绝，对网络信息污染状况做出科学的评价，将此问题置于公众的监督之下，有助于提升解决问题的效率。[方法 / 过程] 本文结合 12321 中心数据和公众问卷调研数据，运用模糊语言改进的“拉开档次”方法确定指标权重，使用 TOPSIS 方法集结评价信息。[结果 / 结论]2014 年 2 月至 2016 年 10 月期间，网络信息污染整体状况形势较为严峻，尤以 2015 年 3、4 月时段信息污染最为严重。本文基于模糊语言的改进“拉开档次”方法相比传统“拉开档次”方法，具有更好的评价单元区分能力。

关键词：网络信息污染 拉开档次法 模糊语言评价 12321 中心 社会调查

分类号：G203

引用格式：万晓榆，王在宇，蒋婷．基于“拉开档次”法的网络信息污染状况模糊语言评价研究 [J/OL]. 知识管理论坛，2017, 2(2): 145-155[引用日期]. http://www.kmf.ac.cn/p/1/121/.

互联网是一把双刃剑，在带给人们便利的同时也带来了网络信息污染现象，如垃圾短信、骚扰电话、垃圾邮件等，造成了许多负面的影响。12321 互联网不良与垃圾信息举报中心（以下简称 12321 中心）数据显示，最近一年我国网民因为网络信息污染所遭受的经济损失高达 915 亿元人民币，几乎接近 2015 年西藏地区全年的 GDP 总和（2015 年西藏地区 GDP 总和为 1 026 亿元）；给全体网民造成的时间损失相当于 3 822 人的生命（2015 年世界卫生组织报告中国人均寿命，男性 74 岁，女性 77 岁，按平均寿命 75 岁计算）。那么，当前网络信息污染状况到底呈现怎样的状态？公众对网络信息污染的认知与中心统计数据是否一致？如何对网络空间中的信息污染状况进行科学评价？通过回答这些问题，并将网络信息污染问题置于公众的监督之下，有助于营造“天气清朗、生态良好”的网络空间环境。

信息污染一词由德国学者拉斐尔普罗在论文《信息生态学进展》中首次提出 [1]。信息污染主要包含两类：一是媒介信息中混入的有害、有毒、具有欺骗性和误导性的信息；二是过载信

基金项目：本文系国家社会科学基金项目“新兴信息消费的环境评估及治理研究”（项目编号：15XGL024）研究成果之一。
作者简介：万晓榆 (ORCID：0000-0001-7875-7175)，院长，教授，博士；王在宇 (ORCID：0000-0003-2159-9975)，硕士研究生，通讯作者，E-mail：zaiyuw@outlook.com；蒋婷 (ORCID：0000-0002-8805-7516)，本科生。
收稿日期：2017-01-05 发表日期：2017-04-12 本文责任编辑：王传清



息[2]，这些污染信息可能会导致人们隐私的泄漏以及时间、金钱和精神层面的损失。随着大数据、云计算、移动互联网等新兴信息技术的应用，信息污染已经成为网络治理和舆论引导研究领域的热点问题[3]。现有研究主要从以下两个方面展开。

一方面是对信息过载问题及其解决方案的研究。信息过载是指人们接受了过多信息，但却无法有效整合、组织及内化成自己需要的信息，以致影响到工作、学习和人际关系[4]。M. J. Eppler 等将信息过载产生的原因分为个人因素、信息特质、任务及过程因素、组织设计及信息科技五大类别[5]。何仲等认为信息过载问题在网络购物环境中会让消费者选择困难，造成消费者时间和心理上的成本浪费[6]。王娜和郑巧伟研究微信订阅服务中出现的信息过载现象，发现当前微信订阅服务的内容存在重复、低质等问题[7]。那么信息过载问题如何解决呢？大多数学者选择从技术角度降低数据维度，帮助用户获取有用信息，主要包括个性化推荐算法和搜索引擎技术。梁劳慧从信息组织者出发，认为图书馆应该通过对用户开展信息素质教育、制订咨询手册等措施帮助用户避免信息过载现象[8]。王娜等对泛在网络中的信息过载问题进行了抽样调查和分析，并提出了在移动社交网站中建立个性化推荐的机制[9, 10]。王娜和田晓蒙研究了豆瓣社区中信息的组织设计对信息过载产生的影响，并针对豆瓣平台提出了大众分类法优化方案[11]。王又然基于加权小世界网络理论，发现以人人网为代表的社交网络站点中单个社群内成员之间特征关系长度过短是造成该类平台出现重复率高、同质化高的原因[12]。高锡荣等通过从知网搜索关键词，建立信息价值评估指标体系[13]。赵静娴将在线评论分为非垃圾评论、欺骗性评论、干扰性评论和低效用评论 4 个类别，建立了垃圾评论特征属性表，并基于神经网络和决策树方法设计了在线评论文本分类方法[14]。

另一个方面是对网络信息安全和个人隐私风险评估方面的研究。夏日首次提出了信息污染指数（information pollution index, IPI）的概念[15]，并建立了一套指标体系[16-17]，但他对信息污染源的划分包括实物型、文献型、电子型和网络型，指标包含过多冗余信息，已经不符合网络社会的实际情况。程艳林提出了网络信息污染检测的 7 个维度[18]，但没有给出具体的量化方法。谢友宁和钮钦调研了农民生活中存在的生活型和生产型信息污染[19]，但存在主观性过强的缺点。陈桂香则调研了大学生遭遇信息污染现状及态度[20]，但仅限于对结果的统计描述。T. H. HSU 等运用 ANP 方法对旅游网站进行了评估，发现安全”是人们最关注的问题[21]。朱光等基于信息系统的安全模型[22]，将隐私风险因素分为平台脆弱因素、用户行为脆弱因素和外部威胁因素，构建了社交网络隐私风险评估体系，并用模糊评价方法进行评价[23]。

通过对信息污染相关研究的总结，发现当前的研究主要聚焦于技术层面。实际上信息污染的源头是人类自身。随着移动互联网的普及，信息污染的影响范围也逐步扩大，但互联网信息污染防治是一个社会软系统问题，具有模糊性、复杂性、系统性等特点[24]，因此从技术角度无法从本质上解决问题。社会学研究表明，将问题置于公众的评价之下有助于提升人们解决问题的效率，故从公众角度研究信息污染的负面影响，对于警醒人类反思自身的信息行为，提升信息素养，共同维护一个良好的网络空间环境具有重要意义。在信息污染及其影响评估方面，少数学者进行了一定程度的探索，但存在指标体系过于宽泛，不符合当前网络社会实际等问题。

笔者结合 12321 中心提供的信息污染及垃圾信息统计数据和公众遭遇信息污染的态度调查问卷，运用基于模糊区间可能度算法构建对称矩阵，并基于该对称矩阵获取主观权重，再运用“拉开档次”法进行二次加权处理，对传统“拉开档次”法“毫无主观色彩”的权重结果进行修正。从结果对比来看，笔者设计的方法可以较为全

知识管理论坛，2017(2):145–155
DOI：10.13266/j.issn.2095–5472.2017.016

面地评价网络信息污染状况。

1 网络信息污染状况评价指标体系构建

信息污染包含两大类污染源，但考虑到信息过载具有主观性强的特点，不同的人有不同的感受，目前也缺乏相应的统计支撑。故笔者主要针对第一种类型污染信息进行研究，暂不考虑第二类信息污染。

基于前期研究成果[2]，笔者将 12321 中心统计类目作为网络信息污染的 7 类衡量指标，即垃圾邮件、不良网站、垃圾类短信、涉嫌违法类短信、垃圾彩信、骚扰电话和手机应用安全问题（APP），这是目前为止对网络污染信息的最为详细和准确的统计。

同时，在前期调研发现，12321 中心的统计数据与公众的感知存在较大的差别，例如恶意 APP 在统计数据中大量存在，甚至超过了垃圾短信和骚扰电话的举报数量，但公众反映的情况却刚好相反。因此，为了弥补统计数据的片面性缺点，借鉴故障检测领域失效模式及影响分析（failure mode and effects analysis，FMEA）方法[25-26]中采用发生度、严重度和检出度对失效模式进行风险评估，笔者针对现有统计的 7 种污染信息类型，设计了《公众遭遇信息污染的态度调查》问卷，从公众遭遇信息污染的频率（是否经常遇到）、已对公众造成的危害程度（心理伤害、时间浪费或钱财损失）以及公众对各类污染信息的厌恶程度这 3 个角度进行公众认知态度调查。以此综合考虑统计数据和公众认知态度，全面评价网络信息污染状况。

2 网络信息污染状况评价模型与方法

在建立评价指标体系之后，就是获取指标权重和选择信息集结方法。研究路线如图 1 所示。

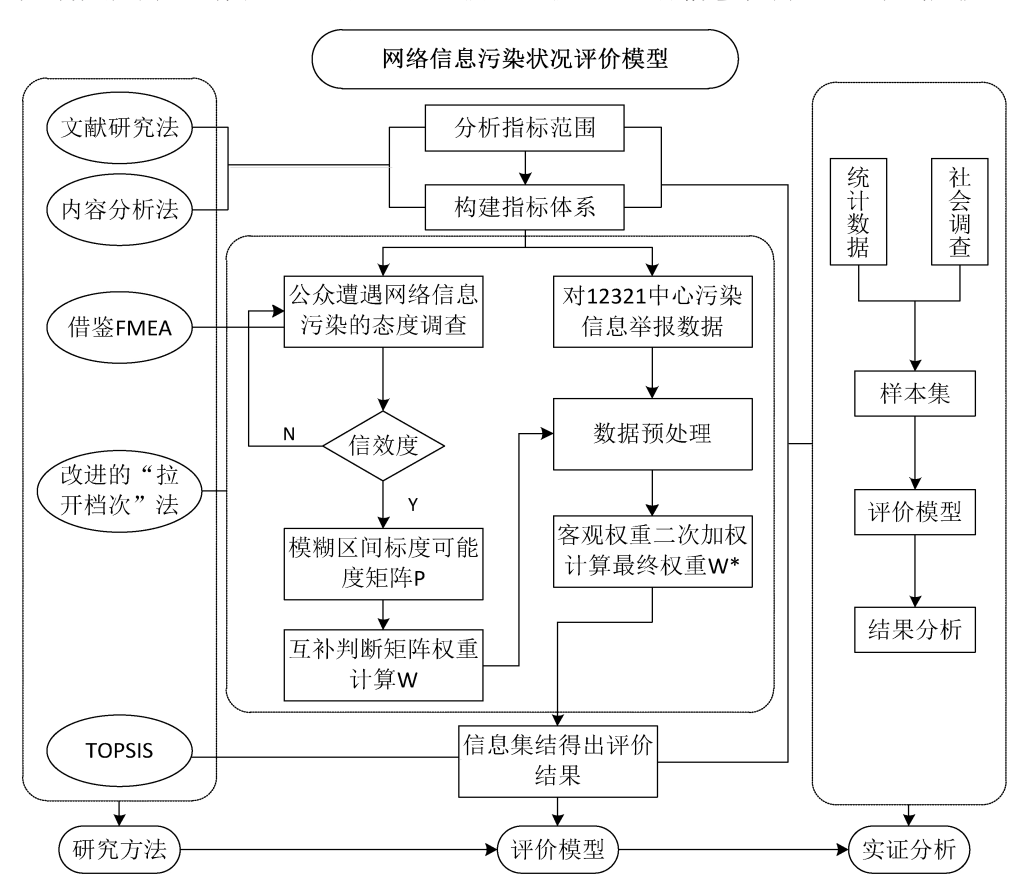


图 1 网络信息污染状况评价研究路线



知识管理论坛
2017 年 第 2 期（总第 8 期）

2.1 基于模糊区间可能度的互补矩阵权重获取方法

由于人们更喜欢使用语言对事物进行评判，如“优”“良”“中”“差”，而人类语言具有模糊性和复杂性的特点，所以学者们基于模糊数学提出了相应的权重获取方法。徐泽水提出了模糊语言标度的区间可能度方法 [27]，可以从原始数据中获取权重。笔者将这种方法应用于评估矩阵是模糊语言的情形。

2.1.1 模糊语言标度的区间可能度

定义 1 设矩阵 $A=(a_{ij})_{n\times n}$，若有 $0\le a_{ij}\le 1$，则称矩阵 A 是模糊矩阵。

定义 2 设模糊矩阵 $A=(a_{ij})_{n\times n}$，若有 $a_{ij}+a_{ji}=1$，则称矩阵 A 是模糊互补矩阵。

定义 3[28-30] 设区间数 $a=[a^-,a^+]$，$b=[b^-,b^+]$，定义区间数的运算法则：

1）间数的加法：$a+b=[a^-+b^-,a^++b^+]$；

2）数与区间数的乘法：$\lambda a=[\lambda a^-,\lambda a^+]$，其中 $\lambda\ge 0$；

3）若 $a^-<b^-,a^+<b^+$，则称 $a<b$。

定义 4[27] 设区间数 $a=[a^-,a^+],b=[b^-,b^+]$，且记 $l_a=a^+-a^-,l_b=b^+-b^-$，记 $a\ge b$ 的可能度公式如公式（1）所示：

$$p(a\ge b)=\max\left\{1-\max\left\{\frac{b^+-a^-}{l_a+l_b},0\right\},0\right\}\quad 公式(1)$$

$p(a\ge b)$ 具有互补性，即：$0\le p(a\ge b)\le 1, p(a\ge b)+p(b\ge a)=1$。

2.1.2 基于模糊区间标度可能度的互补判断矩阵权重获取方法

设被调研对象 $d_k\in D$，其中 $D=\{d_1,d_2,\cdots,d_t\}$ 对污染信息类型 x_i 在属性 $G_j\in G$={遭遇信息污染的频率，已受到的危害程度，对信息污染的厌恶程度}下的模糊语言评估值构成的矩阵 $R=(r_{ij})_{n\times m}$，且 $r_{ij}\in S$，其中

S={非常低，有点低，一般，比较高，非常高}为模糊语言标度，与其对应的区间数表达式进行转换 [27]：

非常低 =[0,0.2]，有点低 [0.2,0.4]，一般 =[0.4,0.6]，有点高 =[0.6,0.8]，非常高 =[0.8,1] 基于模糊区间标度可能度的主观权重的过程如下：

步骤 1：对所有调研对象给出的模糊区间矩阵按公式（2）集结，其中 r_{ij} 通过模糊语言标度进行转换：

$$r_{ij}^*=\frac{1}{t}\sum_{k=1}^{t}r_{ij}\quad 公式(2)$$

为表达方便，集结后的模糊区间矩阵仍记为 $R=(r_{ij}^*)_{n\times m}$；

步骤 2：对污染信息类型 x_i 的各属性值进行集结，$r_i=\sum_{j=1}^{m}r_{ij}^*$；

步骤 3：利用公式（1）对 $r_i,i=1,2,\cdots,n$ 进行两两比较得可能度矩阵 $P=(p_{ij})_{n\times n}$，由定义可知，可能度矩阵为互补判断矩阵，根据公式（3）计算排序权重 [27, 31]

$$\omega_i=\frac{1}{n(n-1)}\left(\sum_{j=1}^{n}p_{ij}+\frac{n}{2}-1\right),i=1,2,\cdots,n\quad 公式(3)$$

2.2 运用“拉开档次”法进行二次加权

在郭亚军提出的“拉开档次”法中，评价指标的权重不再体现指标间的相对重要程度，而是将各指标以同等“地位参与”评价过程，该方法“毫无主观色彩”[32]。但事实上，各项指标相对于评价目标的重要性程度往往不相等。因此，首先利用 2.1 部分得出的主观权重原始数据进行加权，然后再采用“拉开档次法”二次加权。计算步骤如下：

步骤 1：指标一致化和无量纲化处理。笔者使用标准差标准化处理方法，计算如公式（4）所示：

$$x_{ij}^*=\frac{x_{ij}-\bar{x}_j}{s_j}\quad (i=1,2,\cdots,n;j=1,2,\cdots,m)\quad 公式$$

其中，$\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}$，$s_j=\frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2,(j=1,2,\cdots,m)$ 分别为样本均值和标准差；

步骤 2：使用公式（3）得到的各指标权重 $\omega_i,(i=1,2,\cdots,n)$ 对无量纲化后的 $(x_{ij}^*)_{n\times m}$ 进行加权 [33]。为表达方便，称加权后的数据仍为

知识管理论坛，2017(2):145–155
DOI：10.13266/j.issn.2095–5472.2017.016

$A=(x_{ij})_{n\times m}$；

步骤 3：运用“拉开档次法”进行二次加权。

首先，构造实对称矩阵，如公式（5）所示：

$$H=A^{T}A \qquad \text{公式 (5)}$$

然后，取 ω^{*} 为 H 的最大特征值所对应的标准特征向量，ω^{*} 即为所求权重向量 [34]。

2.3 运用 TOPSIS 方法进行信息集结

TOPSIS 方法是多目标决策中一种常用的集结评价信息的方法。定义被评价单元 $(x_{i1},x_{i2},\cdots,x_{im})$ 与正理想单元 $(x_{1}^{*},x_{2}^{*},\cdots,x_{m}^{*})$ 之间的加权欧式距离为：

$$y_{i}^{*}=\sum_{j=1}^{m}\omega_{j}(x_{ij}-x_{j}^{*})^{2},i=1,2,\cdots,m \qquad \text{公式 (6)}$$

同理，设被评价单元 $(x_{i1},x_{i2},\cdots,x_{im})$ 与负理想单元 $(x_{1}^{0},x_{2}^{0},\cdots,x_{m}^{0})$ 间的加权欧式距离为：

$$y_{i}^{0}=\sum_{j=1}^{m}\omega_{j}(x_{ij}-x_{j}^{0})^{2},i=1,2,\cdots,m \qquad \text{公式 (7)}$$

定义贴近度指标为：

$$h_{i}=\frac{y_{i}^{0}}{y_{i}^{0}+y_{i}^{*}} \qquad \text{公式 (8)}$$

显然，贴近度 h_i 越大越好，此时被评价单元与负理想单元距离较大，而与正理想单元距离较小。

3 网络信息污染状况评价

3.1 公众遭遇信息污染的态度调查

问卷发放前期，对问卷内容经过反复讨论，修改问卷题目描述。在此基础上，经过 30 人的小样本发放，进一步调整问卷表达，确保问卷题目清晰明了。问卷共包含三部分：①基本信息，包括性别、年龄、职业、周上网时长、手机系统类型；②主体部分，共 21 个题项；③设置开放式问题 1 个，即“您认为还有哪些信息也属于信息污染？它们对您造成了什么样的危害？”。在 2016 年 11 月至 12 月期间，问卷通过“问卷星”平台和重庆地区人群最为集中的解放碑商圈进行随机发放。

线下问卷发放前对调查员进行了必要的培训。在问卷发放过程中，采取一对一的形式，对问卷填写过程进行指导，使其能按照自身真实情况填写，保证结果的可靠性。在专家指导下，共发放 438 份，收回有效问卷 417 份，有效率 95.2%。本次调查涉及的被调查对象中，男性为 189 名，占比 45.3%，女性为 228 名，占比 54.7%。被调研对象的职业、年龄、周上网时长、手机系统类型如图 2- 图 5 所示：

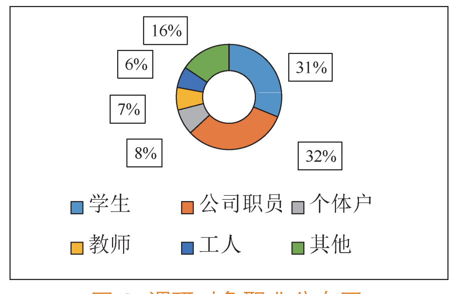


图 2 调研对象职业分布图

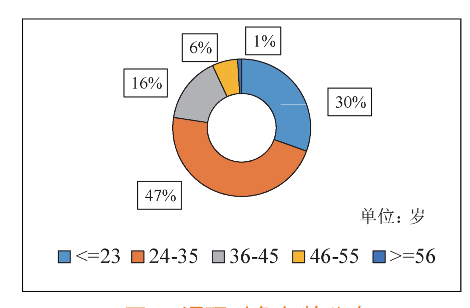


图 3 调研对象年龄公布

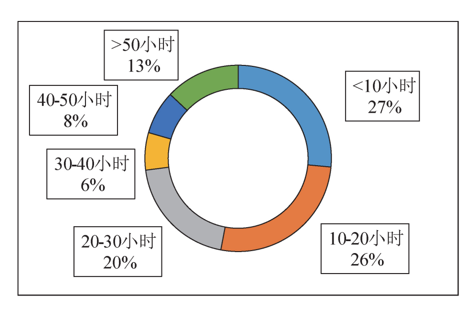


图 4 周上网时长图

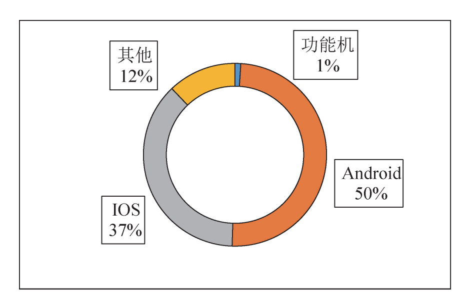


图 5 手机系统类型



在线问卷结果由问卷星自动统计，无缺失数据；线下问卷由两名研究员使用 EpiData3.1 软件背对背录入，以降低录入错误率，此部分问卷存在部分缺失数据且小于 5%，在 SPSS 中采用最大期望值（expectation maximization，EM）方法进行补齐。

3.2 问卷信度和效度分析

3.2.1 信度即可靠度

经 SPSS 计算分析，417 份问卷的信度水平较高（大于 0.7 为高信度），满足探索性研究的要求。各题项的信度系数如表 1 所示。

表 1 问卷信度系数

问卷题项	测量题项	删除后 alpha	alpha
问卷整体信度			0.925
信息污染的遭遇频率	Q1	0.827	
	Q2	0.812	
	Q3	0.814	
	Q4	0.804	0.839
	Q5	0.818	
	Q6	0.825	
	Q7	0.819	
信息污染的危害程度	Q1	0.892	
	Q2	0.889	
	Q3	0.892	
	Q4	0.887	0.908
	Q5	0.894	
	Q6	0.907	
	Q7	0.897	
信息污染的厌恶程度	Q1	0.894	
	Q2	0.891	
	Q3	0.899	
	Q4	0.895	0.911
	Q5	0.896	
	Q6	0.909	
	Q7	0.896	

3.2.2 效度即有效性

常见的问卷效度可以分为内容效度和结构效度，本研究所使用问卷的内容效度经 5 名专家讨论通过。结构效度要求各维度下因子具有一致性，而本研究中所使用的问卷是要测量出公众对不同类型污染信息的认知态度，因此结构效度不适用于本研究目的，故无需作结构效度检验。

3.3 权重计算与比较分析

从 12321 中心获取到的 2014 年 2 月至 2016 年 10 月各类污染信息的统计数据 (如表 2 所示)，然后计算 3 种权重。

3.3.1 客观权重的计算

先将附表中的数据按照公式（4）进行标准化处理，然后运用公式（5）构造的实对称矩阵 H（见表 3），求实对称矩阵 H 最大特征值对应的向量得客观权重，见表 4 中第 2 列。

在 7 类染污信息中，“垃圾彩信”权重最大为 0.174，而骚扰电话权重最小为 0.109，这正好与公众对各类污染信息的反感程度呈反比——对越讨厌的污染分配越小的权重，使其得分越低。这也和我们的直观感受一致，即人们通常会给不喜欢的事物较低的评价，例如人们网购时，如果对收到的网购物品不满意，通常会就给予差评。

客观权重在“垃圾彩信”这类污染信息上赋值较大，而在“恶意 APP”上的赋值较小。这是由于这两类污染信息的举报数量与其他类型的污染信息存在数量级上的差异，为了让各指标地位均等地参与评价，“拉开档次法”对举报数量最多的“恶意 APP”指标分配了较小的权重，而对举报数量最少的‘垃圾彩信’分配了较大的权重。

3.3.2 公众主观权重的计算

对问卷信息集结后计算出的可能度矩阵 P（见表 5）。再由公式（3）计算出公众主观权重 ω，如表 4 中第 4 列所示。

3.3.3 综合权重的计算

将公众主观权重 ω 与标准化矩阵 A 加权后，再利用“拉开档次”法二次加权，得综合权重，如表 4 中第 6 列所示。与传统“拉开档次”法相比，基于统计数据得出的客观权重得到了修

ChinaXiv合作期刊
知识管理论坛，2017(2):145-155
DOI：10.13266/j.issn.2095-5472.2017.016

表 2 2012 年 2 月 -2016 年 10 月 12321 中心垃圾信息举报统计

时间段	B1	B2	B3	B4	B5	B6	B7
201402	4 875	6 278	3 713	4 154	369	3 647	351 311
201403	8 229	10 953	7 156	9 914	628	10 387	316 318
201404	7 801	20 891	6 303	9 972	462	11 640	341 323
201406	7 559	22 532	8 208	10 504	403	13 546	303 995
201407	11 151	25 058	14 219	11 919	317	16 431	302 723
201408	10 563	27 236	12 035	8 761	170	15 241	345 079
201409	9 879	30 238	8 242	8 280	181	22 476	279 247
201410	5 010	24 075	9 655	10 061	279	27 433	232 123
201411	7 976	22 349	13 440	10 183	372	29 614	132 777
201412	7 369	21 473	16 157	11 682	188	34 068	94 489
201501	6 946	26 409	14 059	11 743	230	33 290	78 826
201502	5 384	23 152	8 475	3 677	94	12 409	94 930
20150304	38 528	61 558	23 074	16 477	442	65 574	169 243
201505	24 101	34 035	13 090	8 317	166	31 882	66 762
20150607	8 550	75 541	28 490	25 993	577	72 162	54 143
201508	7 902	34 035	9 648	11 644	233	38 411	23 728
201509	9 438	28 407	9 455	13 026	137	36 694	35 001
201510	10 775	30 927	7 923	14 262	163	39 777	24 737
201511	11 667	25 989	7 703	11 090	158	47 914	22 294
201601	7 746	29 152	6 717	10 523	161	37 255	33 332
201602	4 890	20 430	2 945	3 516	62	9 149	30 040
20160506	17 415	54 164	10 906	18 662	463	69 900	90 240
201607	6 121	30 529	8 770	13 512	221	33 957	140 221
201608	8 123	31 151	7 989	12 162	201	41 729	145 697
201609	8 616	20 193	7 475	11 741	126	41 709	64 626
201610	5 639	22 050	7 618	8 058	189	18 766	75 094

注：B1 垃圾邮件（封 / 月），B2 不良网站（个 / 月），B3 垃圾短信（条 / 月），B4 违法短信（条 / 月），B5 垃圾彩信（条 / 月），B6 骚扰电话（起 / 月），B7 恶意 APP（款 / 月）

表 3 仅由统计数据计算所得实对称矩阵

	B1	B2	B3	B4	B5	B6	B7
B1	5.38E-07	1.48E-07	3.43E-07	2.75E-07	5.23E-06	1.06E-07	-1.8E-09
B2	1.48E-07	1.28E-07	2.63E-07	3.35E-07	3.94E-06	8.56E-08	-5.3E-09
B3	3.43E-07	2.63E-07	8.86E-07	7.77E-07	1.32E-05	1.67E-07	-5.2E-09
B4	2.75E-07	3.35E-07	7.77E-07	1.28E-06	1.92E-05	2.71E-07	-1.3E-08
B5	5.23E-06	3.94E-06	1.32E-05	1.92E-05	0.001192	1.8E-06	7.02E-07
B6	1.06E-07	8.56E-08	1.67E-07	2.71E-07	1.8E-06	7.94E-08	-6.9E-09
B7	-1.8E-09	-5.3E-09	-5.2E-09	-1.3E-08	7.02E-07	-6.9E-09	1.95E-09



知识管理论坛
2017 年 第 2 期(总第 8 期)

表 4 三种权重分配结果对比分析

污染信息类型	客观权重	权重排名	公众主观权重 ω	权重排名	综合权重 ω^*	权重排名
垃圾邮件	0.004 236 1	4	0.150 129 6	4	0.004 068 0	4
不良网站	0.003 194 5	5	0.148 047 8	5	0.002 569 6	5
垃圾短信	0.010 701 9	3	0.112 668	6	0.011 202 8	3
违法短信	0.015 534 2	2	0.153 208 8	2	0.015 335 9	2
垃圾彩信	0.964 305 6	1	0.174 132 6	1	0.964 224 5	1
骚扰电话	0.001 460 1	6	0.109 246 6	7	0.002 040 7	6
恶意 APP	0.000 567 5	7	0.152 566 5	3	0.000 558 5	7

表 5 公众问卷信息模糊区间标度可能度矩阵

	V1	V2	V3	V4	V5	V6	V7
1	0.5	0.51247	0.7247	0.481564	0.356085	0.745207	0.485416
2	0.48753	0.5	0.712145	0.469091	0.343652	0.732648	0.472942
3	0.2753	0.287855	0.5	0.256691	0.131082	0.520588	0.260542
4	0.518437	0.53091	0.743309	0.5	0.374436	0.763825	0.503856
5	0.643915	0.656348	0.868918	0.625565	0.5	0.889399	0.629424
6	0.254794	0.267352	0.479412	0.236175	0.110601	0.5	0.240025
7	0.514584	0.527058	0.739458	0.496144	0.370576	0.759975	0.5

正，使得公众最讨厌的骚扰电话和垃圾短信权重略有增加，公众诟病较少的污染信息类型的权重略有降低。除“垃圾短信”和“恶意 APP”的权重分配差异较大外，其余类型的污染信息所分得的权重比较接近。即公众对这些污染信息的直观感受和 12321 中心的统计数据较为一致。而权重差异较大的两类污染信息是垃圾短信和恶意 APP。对于垃圾短信，公众反感程度很强烈，而在统计数据上可能还存在未统计到的部分。通过访谈得知，超过五成的公众在对垃圾信息的处理方式是直接删除，而未选择举报。有些被调查者的手机上虽然安装了手机卫士，具备拦截功能，但是垃圾短信在被拦截时，因为担心可能有短信被误操作，所以仍然会选择查看，查看后也一般选择删除相关短信。对于恶意 APP，权重分配差异较大的原因仍然是公众的直观感受和统计数据之间的差异。目前各类手机应用商店都和 12321 中心有合作，用户在下载 APP 以前，大量恶意 APP 已经被过滤掉了，所以公众对此类污染信息感知较少，而在 12321 中心的统计数据中却大量存在。因此，在 7 类污染信息上的权重分配是合理的。

3.4 评价信息集结

由于参与评价的指标均为负向指标（指标值越小越好，表示污染信息越少，也称成本型指标），故设置正理想系统为各指标标准化后的最小值，设置负理想系统为各指标标准化后的最大值。最后，用 2.3 节中的 TOPSIS 方法集结评价信息，并将其放大 100 倍，使得分在区间 [0,100] 内，更加符合日常习惯。得分越高表示网络信息污染越少，反之，网络信息污染越严重。基于改进后的“拉开档次”法的最终结果与运用传统“拉开档次”法计算客观权重得出的评价结果对比如表 6 所示：

表 6 左侧是基于改进后“拉开档次”法所得出的评价结果。在评价时间段内，从整体上看，多数时段网络信息污染状况评分集中在 80 分左右，情况不是十分理想。2016 年 1 月和 2016 年 8 月评分较高，网络信息污染相对较少，网络信息环境较好；得分最低的是 2015 年 3、4 月，得

知识管理论坛，2017(2):145–155
DOI：10.13266/j.issn.2095–5472.2017.016

分仅 14.84 分，其原因主要是在这一时段内的信息污染可能较为集中，而非时间跨度较大。这是因为 2016 年 5、6 月和 2015 年 6、7 月同样时间跨度是两个月，得分却也与平均水平相差不大，反倒是 2015 年 3、4 月得分明显偏低。

表 6 改进“拉开档”法与传统方法的评价结果比较

时间段	改进法得分	传统法得分
201601	98.80	95.70
201408	98.22	94.73
201610	80.77	75.62
201402	80.75	75.59
201403	80.73	75.60
201502	80.67	75.56
201607	80.58	75.44
201602	80.46	75.32
201508	80.38	75.25
201406	80.34	75.20
201410	80.26	75.15
201509	80.22	75.11
201412	80.20	75.09
201608	80.18	75.07
201409	80.15	75.04
201404	80.05	74.94
201505	80.02	74.92
201407	80.01	74.91
201501	79.98	74.89
201609	79.95	74.84
201511	79.91	74.83
201411	79.49	74.43
20160506	79.37	74.33
20150607	79.31	74.27
201510	79.11	74.08
20150304	14.84	19.33

表 6 右侧是基于传统“拉开档次”法对统计数据进行最终排名的结果。从两种评价结果的对比可以发现，改进后的方法得出的结果评分区间范围由 [19, 96] 扩大到 [15,99]。整体排名差异不大，而 2014 年 2 月和 2014 年 3 月的信息污染状况排名刚好交换了顺序。好的愈好，差的愈差，改进后的方法相当于对被评价单元作了极化处理，这样能够更加明显地展现出各时间段内的信息污染状况。可见，改进后的算法对被评价单元具有更好的区分能力。

4 结论与展望

本文设计了公众遭遇信息污染的态度调查问卷，基于模糊区间标度可能度对称矩阵计算公众主观权重，再运用“拉开档次”法进行二次加权，可以实现对由统计数据计算出的客观权重的修正。运用改进后的“拉开档次”法对 2014 年 2 月至 2016 年 10 月间的信息污染状况进行了评价。与传统方法相比，改进的方法扩大了评分区间，而且 2014 年 2 月和 3 月的评分表现刚好交换了位置。据此可以认为改进后的方法对评价单元具有更好的区分能力，更好地体现“拉开档次”的特点。

相较以往研究，本文将公众态度调查与统计数据相结合，更加全面地衡量了当前的信息污染状况。整体而言，信息污染形势相当严峻。运用改进后的“拉开档次”法，多数时间段的评分在 80 分左右，而以 2015 年 3、4 月的信息污染最为严重。同时，问卷调查结果发现，公众对 7 类污染信息的感知与中心的统计数据间存在较大差异。如中心统计数据显示，“恶意 APP”举报数量显著高于其他类型的污染信息，但我们在社会调查的过程中发现，公众对“垃圾短信”和“骚扰电话”最为反感，遭遇恶意 APP 的频率和因其而受到的损失则相对较低。造成这一现象的原因，是公众感知与统计数据间存在差异。在“恶意 APP”方面，可能是手机安全卫士软件以及各大应用商店的相关审核措施起到了效果；而在垃圾短信方面，近五成公众接收到垃圾信息后存在直接删除的情况，导致部分垃圾短信无法进入 12321 中心统计数据；而公众



知识管理论坛
2017 年 第 2 期（总第 8 期）

对骚扰电话最为反感的原因，可能是因为电话属于媒介丰富度较高的沟通形式，因此骚扰电话给公众造成的主观印象最深，通过回答问卷的形式得到了反映。此外，“垃圾彩信”在中心统计数据和公众反映中都处于较低水平。随着QQ、微信等即时通讯的发展，彩信正在成为历史，因此建议将彩信与短信类污染信息合并统计即可。

本文也存在不足之处。首先，本文主要针对第一类网络信息污染状况进行了评价，但在问卷调查过程中发现以微信为代表的社交工具中的信息过载问题严重，公众反映较为强烈。其次，在问卷调查时，为了能够提高问卷的可靠性，本文选择线上和线下两种渠道回收问卷，其中线上问卷受众涉及全国。但由于资源限制，通过两种渠道所获得的样本中，重庆地区的人群比例较大，地区特征均较为明显。在有条件的情况下，可以考虑在全国范围内进行分层抽样，使人群覆盖更广，更具代表性。最后，在研究中仅对 2016 年 11 月至 12 月期间公众遭遇信息污染的态度进行了调查。虽然从心理学的角度来说，人们对负面信息的印象更深，持续的时间更长，仍建议后续研究可以每年展开一次社会调研，以反映公众认知的动态变化，使评价结果更加具有实效性。

参考文献：

[1] 周庆山 , 李瀚瀛 , 朱建荣 , 等 . 信息生态学研究的概况与术语界定初探 [J]. 图书与情报 , 2006(6):24-29.

[2] 刘雪艳 , 王在宇 , 袁野 . 基于内容分析法的网络信息污染演变研究 [J]. 现代情报 , 2016(7):45-50.

[3] 申楠 , 杨琳 . 复杂背景下网络舆论引导与网络环境治理探析 [J]. 西安交通大学学报 (社会科学版), 2014(4):96-101.

[4] 蔺丰奇 , 刘益 . 信息过载问题研究述评 [J]. 情报理论与实践 , 2007,30(5):710-714.

[5] LIN C. Online stickiness: its antecedents and effect on purchasing intention[J]. Behaviour & information technology, 2007, 26(6):507-516.

[6] 何仲 , 张念照 , 吕廷杰 . 信息过载环境下网络消费者购买意愿的形成研究 [J]. 价格理论与实践 , 2013(4):95-96.

[7] 王娜 , 郑巧伟 . 微信订阅服务中信息过载的调查与防控机制研究 [J]. 现代情报 , 2016(10):47-53.

[8] 梁劳慧 . 信息焦虑与信息超载下的图书馆作用分析 [J]. 图书馆学研究 , 2011(1):27-29.

[9] 王娜 , 任婷 . 移动社交网站中的信息过载与个性化推荐机制研究 [J]. 情报杂志 , 2015(8):190-194.

[10] 王娜 , 陈会敏 . 泛在网络中信息过载危害及原因的调查分析 [J]. 情报理论与实践 , 2014(11):20-25.

[11] 王娜 , 田晓蒙 . 大众分类法对信息过载的影响及优化策略研究——以豆瓣网为例 [J]. 现代情报 , 2016(9):74-81.

[12] 王又然 . 社交网络站点社群信息过载的影响因素研究——加权小世界网络视角的分析 [J]. 情报科学 , 2015(9):76-80.

[13] 高锡荣 , 张红超 . 海量文本信息的价值评估模型及算法——以新浪微博为例 [J]. 情报杂志 , 2016(6):151-155.

[14] 赵静娴 . 网络交易垃圾评论智能识别研究 [J]. 现代情报 , 2016(4):57-61.

[15] 夏日 , 程刚 . 信息污染指标体系的构建研究 [J]. 情报理论与实践 , 2005(6):34-37.

[16] 夏日 . 信息污染测度指标体系的构建研究 [J]. 情报理论与实践 , 2009(11):42-45.

[17] 夏日 , 程刚 . 信息污染源样本数据采集方法研究 [J]. 情报理论与实践 , 2008(2):275-278.

[18] 程艳林 . 网络信息污染程度检测模型初探 [J]. 新闻爱好者 , 2009(18):119-120.

[19] 谢友宁 , 钮钦 . 农村信息化背景下的信息污染灾害风险评估 [J]. 浙江农业科学 , 2013(1):109-112.

[20] 陈桂香 . 大学生遭遇信息污染的现状调查研究 [J]. 教育教学论坛 , 2014(53):103-104.

[21] HSU T H, HUNG L C, TANG J W. A hybrid ANP evaluation model for electronic service quality[J]. Applied soft computing, 2012, 12(1):72-81.

[22] WU Y, FENG G, WANG N, et al. Game of information security investment: impact of attack types and network vulnerability[J]. Expert systems with applications, 2015,42(15):6132-6146.

[23] 朱光 , 丰米宁 , 陈叶 , 等 . 大数据环境下社交网络隐私风险的模糊评估研究 [J]. 情报科学 , 2016(9):94-98.

[24] 郭海滨 , 郑丕谔 . 防治互联网信息污染的知识管理概念模型研究 [J]. 情报杂志 , 2006(10):84-87.

[25] 张悦 , 石超 , 方来华 . 基于 FMEA 和 HAZOP 的综合分析方法及应用研究 [J]. 中国安全生产科学技术 , 2011(7):146-150.

[26] 王沙婷 , 梁工谦 . 面向产品再制造的改进 QFD 和 FMEA 的集成研究 [J]. 软科学 , 2011(5):61-64.

[27] 徐泽水 , 达庆利 . 基于模糊语言评估的多属性决策方法 [J]. 东南大学学报 (自然科学版), 2002(4):656-658.

[28] 白先春 , 贺晋 . 一种基于模糊语言评估的多属性决策方

知识管理论坛，2017(2):145−155
DOI：10.13266/j.issn.2095−5472.2017.016

法 [J]. 运筹与管理 , 2006(2) :50-52.

[29] XU Z S, Da Q L. The uncertain OWA operator[J]. International Journal of Intelligent Systems, 2002, 17(6):569-575.

[30] SENGUPTA A, PAL T K. On comparing interval numbers[J]. European journal of operational research, 2000, 127(1):28-43.

[31] 徐泽水 . 模糊互补判断矩阵排序的一种算法 [J]. 系统工程学报 , 2001(4):311-314.

[32] 郭亚军 . 一种新的动态综合评价方法 [J]. 管理科学学报 , 2002(2): 49-54.

[33] 王蕾 , 杨力 . 基于粗糙集和拉开档次法的煤矿本质安全管理评价模型 [J]. 中国矿业 , 2015(11):16-20.

[34] 郭亚军 . 综合评价理论、方法及应用 [M]. 北京 : 科学出版社 , 2007.

作者贡献说明：

万晓榆：拟定论文框架，修改论文；

王在宇：论文撰写，数据建模与分析；

蒋　婷：问卷发放，数据采集。

Research on the Evaluation Model for Network Pollution Status Based on Fuzzy Linguistic Assessments and the Scatter Degree Method

Wan Xiaoyu, Wang Zaiyu, Jiang Ting

School of Economics & Management, Chongqing University of Posts and Telecommunications, Chongqing 400065

Abstract: [Purpose/significance] Network pollution has been hated heavily by the crowd. It’s essential and helpful to enhance the efficiency of solving this complex problem to expose it through a scientific evaluation way. **[Method/process]** Based on 417 questionnaires, a way of gaining subjective weights combined with fuzzy linguistic assessments was designed in this paper. In the end, the TOPSIS method was applied to gather the evaluation results. **[Result/conclusion]** From February 2014 to October 2016, the network pollution status was severe overall. The status of network in March & April 2015 was the worst during the whole time. And the improved scatter degree method in this paper can distinguish the units easier compared with the traditional one.

Keywords: network information pollution　scatter degree method　fuzzy linguistic assessments　12321 center　social survey